Supplementary Appendix

(To accompany: Advancing methods for reliably assessing motivational interviewing fidelity using the Motivational Interviewing Skills Code)

In this Appendix, we describe some aspects of motivational interviewing that can only be assessed with utterance-level coding utilizing the MISC. We provide some additional methodological detail on the process of coding polysubstance abuse using multiple target behaviors and a confusion matrices for commonly confused codes.

Coding MI Sessions with Multiple Target Behaviors

MISC coding has traditionally focused on rating patient change-talk relative to a single, desired target of behavior change (e.g., reducing alcohol use). The MISC manual describes how a coding team should be assigned a list of target behaviors by the principal investigator "so that he or she will know which particular instances of client language to attend to and which to ignore" (Miller et al., 2008, pg. 32). Instances of patient change-talk focused on other outcomes (e.g., reducing marijuana use) are often given the code Follow/Neutral, a generic code assigned to utterances when there is no clear link between the statement and moving towards or away from the target behavior. The problem with a narrowly defined target behavior list is that the coding may miss important clinical information about multiple target behaviors. This can occur when clients may favor reducing use of one substance by substituting another (i.e., producing a statement that indicates change talk for one behavior and sustain talk for another behavior such as "I could cut down pain killers by using marijuana instead") or when an unexpected and clinically relevant target behavior not on the predefined list emerges in the interview

(e.g., if a client begins to talk about reasons to reduce marijuana use, a behavior which is not on the pre-identified target behavior list, these statements would be coded as Follow/Neutral). In this example, coding change talk about marijuana as Follow/Neutral would prevent future analysis examining how these change statements might affect other behaviors like alcohol. The client's speech about behavior change would be essentially lost in the coding by being grouped into Follow/Neutral.

Missing or miscoding change talk limits research that could further inform MI and theories of change. Recently MI theorists have broadened the conceptualization of change within a motivational interview to factors beyond individual target behavior change and have defined behavior change to encompass other aspects of human experience (Miller & Rollnick, 2013). This shift follows common factors research where relational elements like accurate empathy may account for both increased change statements and behavior change (Gaume et al., 2008; Norcross & Wampold, 2011). Nuanced hypotheses about the relationship between client and therapist speech and change are difficult to study when limited to target behaviors that may exclude some forms of change talk as Follow/Neutral language.

Coding of multiple target behaviors can also provide a metric for clinician feedback on focusing the MI interview towards specific target behaviors. This is especially relevant for real-world settings with clients with polysubstance abuse. Where typical narrowly defined assessments artificially remove coding of language about other substances of abuse that may be mentioned, coding client statements relevant to several substances of abuse can provide an accurate measure of the way the clinician and client

shifted from substance to substance or focused on an individual substance target behavior during the interview.

Methods for coding multiple target behaviors. Part of our sample came from a real-world primary care implementation study (Krupski et al., 2012) in which patients were screened for drug problems were randomly assigned to receive a brief motivational intervention or care as usual. The vast majority of the patients from this study reported polysubstance use, and thus, we could not predict how many or what type of substance abuse target behaviors would emerge in a given session. It was ultimately up to the MI counselor to identify the target behavior for the interview when presented with polysubstance abuse. In this trial, most of the counselors were primary care clinic social workers who varied in MI proficiency and skill in focusing the interview on one target behavior. Interviews often began with focus on one target behavior (e.g., alcohol) and moved to another (e.g., alcohol and marijuana) or sometimes focused on the relationship between two or more substances of abuse and thus targeted multiple substances at the same; this was the reality of performing MI in a real-world primary care setting with a polysubstance abusing population. Because our study involved secondary data analysis of this trial and other similar trials with two or more substances, we chose to adjust the coding scheme to the real world data and allow for concurrent coding of multiple target behaviors.

To code multiple target behaviors in our study, coders defined the target behaviors (i.e., substances of abuse) as they emerged in the interview. After identifying the client behavior code (i.e., type of change or sustain talk), they tagged each behavior code with a number that corresponded to a list of potential target behaviors.

We utilized the following numeric coding scheme for the list target behaviors:

1 = Alcohol

2 = Marijuana

3 = Cocaine or crack

4 = Opiates

5 = Methamphetamine or amphetamines

6 = Tobacco

7 = Other Drug 1

8 = Other Drug 2

Numbers 7 and 8 referred to write-in categories for drugs that were not mentioned in the primary list (e.g., hallucinogens, PCP, ecstasy). Drug 7 was first used and if another drug that was not in the list emerged after coding Drug 7, then Drug 8 would also be used. This could continue indefinitely for the number of additional drugs mentioned (e.g., Drug 9, 10, 11); however, in our study we found that two other drug categories was normally sufficient. Other researchers may wish to define their target behavior list according to the most frequently occurring substances of abuse reported in the sample.

There were situations that occurred when a target behavior was discussed but never explicitly named. For instance, the patient referred to their drug of choice as "my drug" or "drugs" and the interviewer did not or could not elicit the specific type of drug discussed. If there was a reference to a single unknown drug the coders wrote "unknown drug" for Drug 7/8. If there was a reference to multiple unknown drugs the coders wrote "polysubstance: unknown."

When two drugs were referenced jointly in client change or sustain talk, the

coders were directed to use both numbers (one for each drug) to double or stack code the utterance. When an utterance was double coded, it was treated as a singular utterance that met criteria for two codes (i.e., neither code received precedence). For example, in a session that focused on alcohol and marijuana, the target behavior codes of 1 and 2 would be used respectively for change talk focusing on each drug. When a statement such as "I need to quit everything" occurred then the utterance would be double coded with Reason+desire1 and Reason+desire2, where the codes would be stacked (e.g., "I need to quit everything" [R+d1;R+d2]). There were situations where a client utterance was double coded in two different directions because the statement indicated change talk for one target behavior and sustain talk for another. For example, "I will cut down on oxycodone by using marijuana for pain instead" would be coded Commitment+4 and Commitment-2.

Some interviews contained mention of three or more substances of abuse that were consistently referenced as if they were one substance (e.g., client statements like "when I use" that referred to a set of different substances of abuse that were mentioned throughout the interview). In these interviews, client did not seem to differentiate usage of different drugs but referred to them jointly and often the counselor did not elicit ambivalence or discuss each drug individually but targeted the polysubstance use like a single behavior. In these situations, the coders were directed to treat the polysubstance use as a single target behavior (i.e., where the target behavior is all substances of abuse). The coders were directed to utilize Drug 7 or 8 write-in category and write in "polysubstance: [target behavior numbers from list, e.g., 1,3,4,6]."

Assessment of reliability for coding multiple target behaviors. Although we described potential benefits of polysubstance coding, we could not assess the individual reliability of behavior codes by target behavior (e.g., Commitment+Alcohol versus Commitment-Marijuana). When divided by substance and behavior code there were not enough utterances (i.e., the cell size was too small) to utilize an intraclass correlation to assess reliability. Given that the overall assessment of utterance-level reliability was adequate to excellent following Cicchetti's (1994) standards, we indirectly assumed that it was feasible to reliably code polysubstance and multiple target behavior change/sustain talk.

Future research on specific versus multiple target behavior coding. Following common factors literature, we speculated that focus on a singular target behavior during the MI interview might be equally as important as creating an environment wherein the client is comfortable talking about one or more types of behavior change. Given the current coding scheme our team and other researchers, should they adopt this coding scheme, may be able to assess if change talk about any behavior, regardless of the particular substance targeted in the interview, may result in change for several behaviors. The data identifying multiple target behaviors in a single interview may also be used to provide feedback to counselors about their directional style (e.g., guiding and following) during an MI interview with polysubstance users (e.g., how did they decide which substance to focus on with the client). Future research could involve developing additional metrics for guiding and following based on tracking individual target behaviors throughout the interview.

Understanding Commonly Confused Codes

MISC data coded at the utterance-level allows for closer inspection of coder disagreement via confusion matrices (Kohavi & Provost, 1998). In predictive models, like a logistic regression, confusion matrices are often used to gauge the accuracy of predictions by comparing predicted versus actual classification of the outcome. In truth, a confusion matrix is nothing more than a cross-classification by two different methods, or in the present case, two different coders. Table 3 shows a confusion matrix based on MISC classifications from two coders for the most common codes in our sample. Agreement is seen in the cells on the matrix diagonal, where both coders have assigned the same code. Cells off the diagonal indicate disagreements between coders. The reliability statistics reported earlier (whether Kappa or ICC) focus simply on whether there is agreement or not, but the confusion matrix allows us to see the nature of the disagreement. For example, when coders disagree about complex reflections (REC), they are most likely to confuse them with simple reflections (RES) and are very unlikely to confuse them with a number of other codes (e.g., AF, ST, SU).

[Insert Table 3 about here]

Confusion matrices are a potential tool to assist trainers in identifying confusable codes and may also suggest potential areas for codebook refinement. Another implication of the data presented in the confusion matrix is that certain codes are more likely to be confused because the individual codes are not wholly distinct. Research has traditionally grouped certain codes along theoretically and empirically informed dimensions (e.g., MI

consistent or inconsistent), but these dimensions do not necessarily follow observed confusability as seen in Table 3. For example, simple and complex reflections are commonly confused but the percentage of complex reflections out of the total reflections per session is used as a summary metric to determine MI adherence.

Appendix References

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290. doi: 10.1037/1040-3590.6.4.284

Gaume, J., Bertholet, N., Faouzi, M., Gmel, G., & Daeppen, J-B. (2012). Does change talk during brief motivational interventions with young men predict change in alcohol use? *Journal of Substance Abuse Treatment*. doi: 10.1016/j.jsat.2012.04.005

Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning, 30*, 271-274.

Krupski, A., Josech, J., Dunn, C., Donovan, D., Bumgardner, K., Lord, S. P., . . . Roy-Byrne, P. (2012). Testing the effects of brief intervention in primary care for problem drug use in a randomized controlled trial: Rationale, design, and methods. *Addiction Science & Clinical Practice, 7*(1), 27. doi:10.1186/1940-0640-7-27

Miller, W. R., Moyers, T. B., Ernst, D. B., & Amrhein, P. C. (2008). *Manual for the Motivational Interviewing Skill Code (MISC), Version 2.1*. New Mexico: Center on Alcoholism, Substance Abuse, and Addictions, The University of New Mexico.

Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change (3rd ed.)*. New York, NY US: Guilford Press.

Norcross, J. C., & Wampold, B. E. (2011). Evidence-Based Therapy Relationships. *Psychotherapy, 48*(1), 98-102. doi: 10.1037/a0022161